Analytical Reasoning in Real World Contexts: Status Memorandum

The ability to solve problems has become more and more relevant in workplace and community. Due to vague definitions, the inherently broad range of tasks, and difficulties of measurement (at least without computer assistance), it was decided that a test for measuring analytical reasoning in real world contexts would be developed as an essential and more easily measurable part of problem-solving ability.

The Institute for Educational Research (IBF) in Germany developed a concept for measuring analytical reasoning in real world contexts for German pupils that has been shown to be a good measure of this ability within that population. This concept, the so-called "project approach" is now being proposed for the Adult Literacy and Lifeskills survey (ALL) (see the framework document immediately following this one).

The main features of this concept are:

- The participant is given a problem that can occur in real life (e.g., "projects" such as buying a bicycle or planning a trip).
- According to the (idealized) sequence of steps, to solve the problem the participant has to work on several independent items concerning the evaluation of objectives, analyzing the situation (gathering information), planning, execution of work-steps, and controlling the results.

Ideas for "project" themes were developed with experts from different countries and eight of those "projects" were further developed for ALL. The selected project themes are:

- Renovating a clubhouse
- Planning a trip and a family reunion
- Setting up a space station
- Buying a bicycle
- Organizing the visit of a foreign chorus
- Organizing a sports meeting
- Looking for a new job
- Looking for a new apartment and moving

In the current phase of development each project includes 8 to 19 items.

The project tasks were tested first in Germany on a sample of about 500 subjects (about 60 per project) to gather data about:

- appeal to adults;
- life relevance;
- appropriate difficulty (for students and non students as well as for people with different educational levels);

- measurability (reliability, difficulty, discriminating power);
- feasibility according to time resources; and
- differences between groups (e.g. gender, age, experience, educational level).

All in all, the results of this study were very positive in regard to the feasibility. None of the drafts have had to be dropped.

Cross-cultural portability data from the test must be gathered from different countries.

A feasibility study will take place in the USA in October and interview studies are currently being conducted in Switzerland with a French version of the projects.

Appropriate translation is a great challenge in this stage of work. To get a translation that is cognitively equivalent to the original version, descriptions of the main features of each task and of the cognitive steps needed to be performed in solving the tasks were developed. Utilizing these descriptions, the translator can check if his or her translation demands the same solving processes.

Up to now this has been done for the German (original) versions of four projects. These will be administered in the US feasibility study. The remaining projects will be tested later this year.

Parallel to the translation work, critical item elements that may determine the difficulty of items were identified, and the items were coded according to these criteria. This process is still underway, as not all project tasks have been coded until now, and to ensure uniformity in coding when working with different persons, the precision of categories must still be improved.

For the interpretation of the results, proficiency levels for analytical reasoning in real world contexts will be identified. Until now the only available results were from the study with German pupils with different projects. Three proficiency levels have been identified based on that data.

The last step before entering the field study will be a profound revision of the projects based on the feasibility studies data and a setting of the scoring procedure. Since the development team started with 8 projects and more items than necessary, by utilizing the results of the feasibility studies it will be possible to get enough projects for measuring analytical reasoning in real world contexts that meet the requirements of the field study.

Analytical Reasoning in Real World Contexts

Executive Summary

Broad consensus exists within the assessment and research communities on the importance of problem solving as a cognitive ability. However, due to the enormous range of tasks and behaviors implied by the term, there is little agreement on the exact definition of problem solving and how it should be measured. In addition, there is also an issue of whether or not problem solving is distinguishable enough from general intelligence/cognition to allow for measurement of a distinct competency.

Therefore the development team has decided to concentrate on analytical reasoning in real world contexts as an essential part of problem-solving abilities. The "project approach" to assessing this ability focuses on the cognitive competencies needed for solving complex tasks. These tasks are embedded in a "project" which itself is constructed in accordance with the individual steps of the problem-solving process. These steps include defining the goal, analyzing the situation, planning the solution, executing the plan, and evaluating the result. Given the situational nature of problem solving, these tasks, or projects, are placed in readily identifiable contexts. The situation, though new, is meant to be conceivable for the participant such that the participant solves tasks that might occur in a similar situation in real life.

"Projects" contain a description of the problem situation, the participant's role in the situation, the participant's task, and a listing of the steps in the work that later items refer to. The items are designed to address the five areas of problem solving outlined above. Multiple-item formats may be used in this approach, including multiple-choice items, items that require respondents to locate and identify requested data, ordering items, correspondence items, and free-response items.

While the "project approach" to analytical reasoning in real world contexts is not designed to yield sub-scale scores, previous experiences with this measurement approach have demonstrated three levels of proficiency for score interpretation: identifying information, ordering and evaluating information, and analyzing dependencies.

Analytical Reasoning in Real World Contexts

The Adult Literacy and Lifeskills survey (ALL) is designed to assess a range of skills thought to be important to social and economic success. Problem solving is one of them and analytical reasoning is an essential subset of problem solving. As we enter the 21st century, the ever fast-changing demands of work require flexible and planned behavior responses. The need for problem solving becomes core to managing changing life and job tasks as autonomously and fully as possible.

The "Project Approach": A Conceptual Framework

Traditional Notions of Problem Solving

Almost 90 years ago, John Dewey described a "problem" as a conflict between goals and available methods. In essence, this characterization has since then been used by psychologists working in different paradigms, although they have differed in the theoretical concepts used to specify what "available methods" are and what constitutes a "conflict." In terms of an information processing approach, for example, "methods" are described as operators that transform a given problem state (within a more or less well defined "problem space") into a new problem state, and a "conflict" arises whenever a goal may not be reached by applying a single operator (i.e., in order to reach the goal, an appropriate sequence of operators has to be searched). In terms of more recent approaches to cognitive psychology, "methods" are units of declarative and procedural knowledge that have to be activated to build and transform mental models. Problem solving may then be characterized as "cognitive processing directed at achieving a goal when no solution method is obvious to the problem solver" (Mayer & Wittrock, 1996).

Such general definitions of "problems" and "problem solving," however, show a lack of specificity. They cover a rather broad range of tasks and behaviors:

- Problem-solving processes may range from spontaneous "insight" to multi-step planning and execution
- Contexts vary from puzzles to authentic tasks
- Domains may be more academic or more practical
- Complexity increases from well-defined reasoning problems to dynamic, intransparent environments
- The scope of problem-solving processes may range from low-level monitoring to single tasks, actions and even to management of long-term projects

Several authors have suggested general models of problem-solving processes, typically involving elements or stages such as

- Define the goal
- Analyze the situation
- Plan the solution
- Execute the plan
- Evaluate the result (see O'Neil, 1997, for a review of the relevant literature).

These process models, however, should best be understood as idealized, normative, and rather superficial models, since experimental psychology has shown that real thinking processes show a much more complicated choreography, leading from explorative to tactical (local) and finally to

strategic (global) solution steps. Benjamin Bloom realized that problem solving involves a mixture of knowledge, understanding, application, transfer, and evaluation. Therefore, he refrained from building some sort of higher-order "problem solving" level into his taxonomy of educational objectives (Bloom & Broder, 1950).

Given this broad variety of problems and solution processes, it is questionable whether the ability to solve certain kinds of problems may generalize within a domain or even beyond single domains to general (and cross-curricular) problem-solving competencies. It also would be difficult to draw clear distinctions between these competencies and related constructs, e.g., the construct of critical thinking, which addresses processes like recognition of assumptions, inference, deduction, interpretation and evaluation of arguments—processes that clearly overlap with steps of problem solving (Ennis, 1996). Moreover, since problem solving involves recombining methods and finding new solution paths for unknown situations, there should be an overlap with the construct of creativity, at least in the sense of flexibility and fluidity of ideas.

The question at hand is whether and how "problem-solving competencies" may be defined as ability dimensions that are empirically identified and distinct from other abilities. Since problem solving seems to be ubiquitous in human cognition, problem-solving competency may be heavily confounded with general intelligence. This issue has been intensively discussed within the European research on "Complex Problem Solving (CPS)" (Frensch & Funke, 1995). The conclusion from a huge number of studies is that, contrary to what was originally hypothesized by Dörner, there are substantial correlations if both constructs are operationalized appropriately: problem solving should be measured in situations that are not totally intransparent (in which case solving behavior would just be some sort of trial and error), and intelligence should be interpreted as "information processing capacity." Süss (1996), reviewing relevant work, concludes that a distinct construct "problem-solving ability" (as operationalized by CPS approaches), is indeed dispensable.

Working within a different paradigm, Sternberg and Kaufman also identify intelligence (at least "successful intelligence" as Sternberg understands it) with problem-solving competency (1998). He defines "analytical" components of intelligence as "identifying the existence of a problem," "defining the nature of a problem," "setting up a strategy," and "monitoring the solution for a problem." In his multi-facet model, "creative" components generate problem-solving options. Finally, "practical abilities," another component of successful intelligence, are needed for applying and implementing problem-solving options.

It follows from these arguments that whatever we regard as a measure of problem-solving competency will be substantially correlated with general intelligence. But, instead of identifying both constructs, it might be useful to focus on the notion of problem solving in a way that problem-solving competency is empirically distinguishable within a hierarchy of cognitive factors as a lower order factor nested within general intelligence. The IBF "project approach" aims at assessing analytical reasoning in real world contexts as such a cognitive factor, an essential part of problem solving ability.

The IBF Focus: Analytical Reasoning as Essential Part of Problem Solving Embedded in Everyday "Projects"

Recent theories of human cognition as well as new developments in assessment methodology stress the role of situational

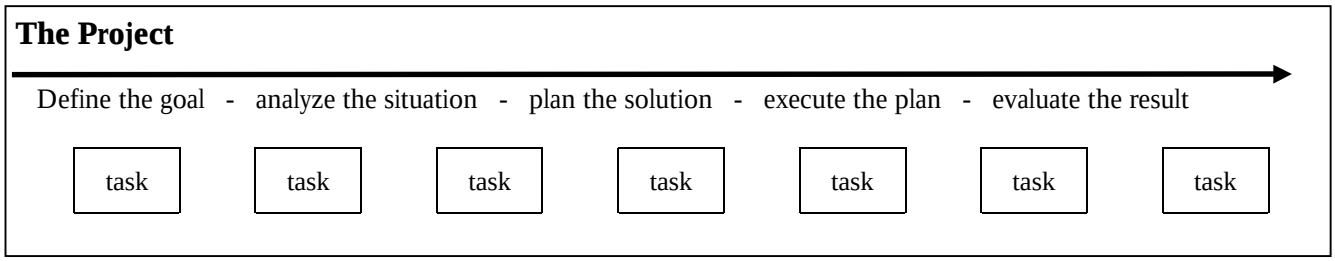
factors in problem solving. Whether or not a "mental tool" (a concept, strategy, or mental model) can be used in a problem situation depends upon features of that situation, such as context, affordances, and constraints. Therefore, test problems which aim at exhibiting the use of mental tools to solve real life problems should be contextualized and made as "authentic" as possible (Mayer & Wittrock, 1996; Cobb & Bowers, 1999).

How can such contextualized, real life problems be defined and worked out as test items? Our answer to this question is grounded in an important body of research on job analysis and vocational training within German applied psychology and educational research. Job analyses for several professional areas came to the conclusion that new forms of labor organization require people to perform beyond routine operations. These "complete actions" include planning, executing, and evaluation steps. Even production workers and office clerks are required to master integrated tasks as described by this "action orientation" approach; the curriculum for vocational training should also be structured according to these tasks. The IBF successfully applied this concept to curriculum development, assessment, and certification reforms in various professions (Hensgen & Blum, 1998; Hensgen & Klieme, 1998). The main idea for assessment is that task training as well as test problems should include all elements of the "complete action."

There are links as well as differences between this assessment approach in vocational training and the IBF approach of assessing analytical reasoning in real world contexts as an essential part of problem solving. A main difference between the assessment of vocational competences based on the model of a complete action and the assessment of problem solving, is that the latter requires, according to the definition of Mayer & Wittrock (1996), situations where no solution method is obvious, and the problem cannot be solved only on the basis of expert knowledge or familiar action-schemes, as it is supposed in vocational assessment. As a strong link, the basic structure of the model of "complete action" is fully compatible with the above-mentioned normative process model for problem solving. Again, it must be stressed that this is an idealized model for problem solving , especially for the factor of analytical reasoning in real world contexts, as well as for a complete action. To describe a more realistic complete action several additional evaluation elements during the whole action should be taken into account. These evaluations may lead to redefinitions of goals and plans.

The "project approach" uses the model of a complete action for test construction. Based on this model, a series of tasks are constructed, which are then integrated into the course of action for an everyday project (e.g., renovating a clubhouse or planning a family reunion). The individual tasks can be solved independently of one another, as both the course of action and the information needed for solving the individual tasks are provided by the test author. Through embedding the individual tasks in an action context, a high degree of context authenticity for analytical reasoning on problem-solving tasks can be obtained. The testee does not solve the problem described at the beginning of a project; rather, he or she solves the individual tasks that may occur in connection with the given action. Figure 1 illustrates the construction of a project, designed as a complete action, including several tasks at different levels of complexity to be solved.

Figure 1: The Project



According to the previously mentioned possible ranges of tasks and behaviors covered by problem solving, the project-approach tasks have the following characteristics:

- well-defined and static
- authentic
- multi-step tasks, which usually need several cognitive steps to find the solution
- embedded in a non-academic domain

Which Psychological Processes are Associated with the Solving of These Tasks?

Project tasks demand analytical operations such as searching, understanding, systemizing, organizing, evaluating, reasoning, and combining information. These cognitive operations are essential for problem solving (defined as information process).

As mentioned above, Mayer & Wittrock (1996) describe problem solving as a "cognitive process directed at achieving a goal when no solution method is obvious to the problem solver." This definition focuses on cognitive processes that "occur within the problem solver's cognitive system and can be inferred indirectly from changes in the problem solver's behavior" (p. 47). As a process, problem solving "involves representing and manipulating knowledge in the problem solver's cognitive system" (p. 47). Project tasks stimulate the analytical reasoning components of problem solving defined as such a cognitive process.

Considerations regarding the definition of the "problem solving" construct are not always limited to cognitive aspects. They also explicitly include active problem solving and the possibility of dynamic problems. An additional problem-solving definition with mention of sub-constructs is offered by Klieme et al. (1998):

"Problem solving is goal-oriented thinking and acting in situations that are new (unknown) to the subject and, therefore, for which no routine solutions exists. Problem solving includes:

1. the systematic gathering, integrating, and structuring of information
2. planning and carrying out steps of action (when dealing with complex, dynamic systems or in experimental thinking, the control of influencing factors through isolation is extremely important)
3. the continual processing of external information and feedback
4. evaluation of one's own actions and their consequences" (p. 2)

Analytical reasoning in everyday contexts, as it is measured by the project tasks, focuses on the sub-constructs of problem solving listed in 1 and 2 above.

There are some limitations of the project approach as some aspects of problem solving that cannot be measured within this approach:

- The "dynamic" aspects of task regulation (continuous processing of incoming information, coping with processes that cannot be influenced directly, coping with feedback, and handling of critical incidents) has to be addressed by computer-simulated tasks (CPS).
- The motivational, affective, and self-regulatory aspects of task regulation might be addressed implicitly by the test tasks or explicitly by questionnaire methods.
- Problem-solving behavior within this test will depend on general, context-specific, domain-specific and situation-specific processes.

Nevertheless, even though not yet proven, it is highly reasonable to assume that a general (latent) competency for analytical reasoning as essential part of problem solving can be identified.

Relations to Other Domains

To what extent does analytical reasoning in real world contexts, when measured with the project approach, overlap or go beyond literacy and numeracy skills? Literacy skills, particularly document literacy and numeracy skills, are necessary for responding to the project tasks. We, therefore, expect that the score will covary with the scores of literacy and numeracy skills. Further, we expect that analytical intelligence in more abstract contexts — not measured in ALL—will influence the score.

However, the project tasks present two special aspects that go beyond literacy tasks and abstract intelligence tasks:

a) They are embedded in a complex, realistic, multi-step action context.

b) They demonstrate a high complexity made possible, for example, through the connection of several sources of information. They therefore require more regulation activity than literacy and numeracy tasks.

The frequently higher complexity of project tasks requires a sufficient understanding of the instructions by the subjects, which itself requires a certain literary competence. This can result in subjects with very low literacy skills not understanding the presentation of the project tasks and therefore being unable to solve them. Literacy skills—according to one hypothesis that will be tested within the framework of ALL—would have a threshold function for the analytical reasoning scale.

Organizing the Domain

As presented in Part 1, the "project approach" for the comprehension of essential subsets of problem-solving competency aims particularly at analytical reasoning in well-defined, contextualized problem situations. The model of problem solving as a complete action serves as a framework for item development and puts analytical reasoning tasks in context.

Therefore, the decomposition of a project into a sequence of steps and questions does not lead to different scores for the various action elements. The rubric for test development, based on this model (complete action with different elements), is purely pragmatic. Reporting will be based on a single latent trait, estimated from data for two or more projects. It is important to combine at least two projects into a problem-solving test, as a certain context-dependency is to be expected for the problem-solving results in the project tasks. Therefore, if at all possible, projects with differing action frameworks should be combined.

For reporting and political considerations, it is important to describe analytical reasoning

competency in real world contexts clearly. For this purpose, we suggest that:

a) after establishment of a one-dimensional IRT-scaling, proficiency levels should be described; and
b) critical item analysis should be carried out, the findings of which indicate which item characteristics contribute to the difficulty of an item.

Guidelines for Test Design Following the "Project Approach"

As has been previously described, the basis for test design is the "model of a complete action" with its five different steps.

Using this model to design test items is, as already mentioned, not new to our Institute. Since the development of the model for assessing vocational action competence, examinations have been developed and applied based on this concept in diverse vocations in both Germany and Luxembourg. The vocational tasks, however, differ in their complexity our "projects," in the assessment methods they use, and in their item formats.

For example, examinations involving a task were developed for the Grand Duchy of Luxembourg. The exam takes about six hours and, in addition to written questions, involves role playing, oral examinations, and completion of practical tasks (e.g., writing a letter using a computer, assembling a device while the candidate's behavior is observed).

At the request of one German state, examinations were developed with no more than four to five written questions for each vocational action but with several actions in an overall test period of about 60 to 90 minutes. Models for practical and oral tests were also developed.

In early 1997, IBF applied the techniques in action-oriented testing to the assessment of analytical reasoning problem-solving abilities. At that time, a test to assess problem-solving abilities for domain-independent and cross-curricula problems was developed on behalf of one of the German states. The test was to be used in the classroom context and administered to pupils about 13 years of age in various types of schools. It was a subtest in a large scale study and consisted of only multiple-choice items. This test was also developed on the basis of the model discussed here and has been successfully tested.

For the purpose of the ALL study we transfer these techniques once again to a different age bracket and different cultural groups. This time, a paper-and-pencil test will be our assessment method that uses different item formats.

How is a Project Structured?

A project will normally have the following elements:

- A description of the problem situation, the candidate's role in this situation, the concrete description of the candidate's task, and a listing of the steps in the work to which the later items refer
- If possible, a separate booklet containing the background information needed to complete the task and solve the items
- Questions referring to the single steps in the work (here there can also be multiple, combined questions) with an introductory description and, if required, additional information in so far as these cannot practically be included in the background information booklet
- An answer key or a sample solution for free-response questions

Which Content Is Addressed in the Steps of the Model?

The following compilation provides a survey, the content of which addresses the

steps in the problem-solving process and, as a consequence, in the items.

Define the Goals	▪ Set goals for yourself. ▪ Recognise which goals are to be reached and identify the essential reasons for the decision. ▪ Recognise which goals/wishes are contradictory and which are compatible. ▪ Assign priorities to goals/wishes.
Analyse the Situation	▪ Select, obtain and evaluate information. - What information is required, what is already available, what is still missing, and what is superfluous? - Where, how, and from whom can you obtain the information? - How are you to interpret the information? ▪ Recognise which people (with what knowledge, tasks and skills) are to be involved in solving the problem. ▪ Select the tools (in the widest sense of the term, including computer programs, vehicles, tools and media) to be used. ▪ Recognise the conditions (e.g., time restrictions, legal conditions) to be taken into account.
Plan the Solution	▪ Recognise which steps need to be taken. ▪ Define the sequence of steps, items on the agenda, etc. ▪ Coordinate work and deadlines. ▪ Make a comparative analysis of alternative plans (recognise which plan is suitable for reaching the goals). ▪ Adapt the plan to changed conditions. ▪ Opt for one plan.
Execute the Plan	▪ Take the individual steps (e.g., write a letter, fill in a form, make calculations).
Evaluate the Results	▪ Assess whether and to what extent the target has been reached. ▪ Recognise mistakes. ▪ Identify reasons for mistakes. ▪ Assess consequences of mistakes.

If one wants to construct items that are independent of one another, then there are two steps—"executing the plan" and "evaluating the results"—that often make for difficulties. Items for these steps (e.g., filling out a form or checking a planned sequence for procedures) will often, though not always, involve steps in planning, situation analysis and even defining goals. If one wants to create a real world context, the fact that several steps in the model will be addressed at the same time within a single item must be accepted.

What Are Possible Questions on the Steps in the Model?

It is possible to derive questions that refer to the different steps from their descriptions above. The following list, which is also available to the test-authors, shows examples of possible general questions.

Define the Goals

- Which goals do you want to reach?

- Which goals should be reached with this action?
- What reasons may have been crucial for this decision?
- Which wishes and demands have to be taken into account?
- Which goals or wishes come into conflict with each other/which goals cannot be reached simultaneously?
- Which goal should be followed in this situation as a matter of priority?

Analyze the Situation

- What information do you need?
- What information is superfluous?
- Which sources of information can you use to answer the following questions?
- How and where can you get the required information?
- What conclusions can you draw from the given information?
- What information should___contain?
- Which persons do you ask for information or assistance?
- What do you need to do the job?
- What kinds of tools, (e.g., graphic, transportation, media, etc.) will be the best?
- What restrictions do you have to take into account?

Plan the Solution

- Which work-steps are necessary?
- Which work-step must be executed next?
- In which sequence must the work-steps be executed?
- Which sequence of events is meaningful?
- Which work-steps must be executed before you can execute this work-step?
- At which date/time can___be done?
- Which goals can be reached easier with the action ___ than with the action ___?
- Which plan meets the demands (or most of the demands)?
- What do you have to do to meet the changed conditions?
- What speaks for Plan A, what for plan B?
- Which plan is more advantageously/more temporally effective/less costly?
- Which plan should be executed?
- What further effects (in the future or side-effects) can the action-plan have?

Execute the Plan

- Calculate, write a letter, fill in the formula, etc.

Evaluate the Results

- Which goal was not reached?
- Which mistakes (concerning....) were made/are in the following document?
- Which causes are responsible for the mistake?
- What consequences will this mistake have?

Which Item Formats Are to Be Used?

Because of feasibility and time-resource concerns in the majority of items, an item type that can be checked by machine is used (these are also short-answer open-response questions). In some cases, however, we use free-response items in order to increase the life relevance of the test.

The following formats are used, but the author is free to develop other formats wherever appropriate. An example for each item type is available in Annex C.

Multiple-Choice Items

In this format, you have to select one or more answers from a number of alternatives, where the number of correct responses may or may not be specified. These alternative responses may take the form of statements, numbers, drawings, document fields identified with letters, sequences of steps, and so on.

Writing Down One or More Numbers or Letters or Combinations of the Two

For example, a price, a flight number, or time of day.

Ordering Items

In this type of item, a certain number of elements are to be put in a proper sequence; for example, work steps must be arranged in the sequence in which they must be carried out.

Correspondence Items

In correspondence items, two series of related elements will be presented, for example, relating information being sought to sources of information. This does not need to be a one-to-one relation; sometimes more than one element can be correctly related. A common example of this item format is the comparison of two alternatives (for example two modes of transportation) where a number of reasons or arguments are presented and the candidate has to decide in each case which of the alternatives supports the use of which mode.

Free-Response Items

The candidate has to write down his or her answers to items that ask for required information, information that is missing, or advantages and disadvantages of certain plans. Filling in formulas or filling in blanks in a text are also possibilities.

What Are the Phases Used in Designing a Project?

The following phases describe the process of test development and some criteria which must be considered.

Phase 1 ***Selecting the Subject for the Project***

This must be a problem that

- is suitable for adults aged 16 to 65 and with varying educational backgrounds
- is relevant to everyday life for the greatest possible number of persons within the target group but is nevertheless a problem
- requires no greater breadth of previous knowledge than that which an entry-level worker would have
- is domain-independent
- is politically acceptable
- can be transferred to other cultures
- has no group-related (e.g., women) bias
- allows for at least six mutually independent items
- covers most of the steps in the process model or, between each item

Phase 2 ***Describing a Problem Situation and Sketching out the Sequence of Actions Necessary to Solve the Problem***

The author's sketch of this sequence of actions should be oriented on the five steps of the process model as well as on the real-world sequence followed in solving the problem.

Phase 3 ***Developing Items on the Steps within the Task***

- Formulating the statement of the problem and the question
- Providing the required information (for inclusion in the background information booklet or accompanying the statement of the problem)
- Selecting a suitable item format
- Developing solutions, distractors, and answer keys or sample solutions for free-response questions

Phase 4 ***Examining and Revising the Project with Its Items***

The authors, at least five other colleagues and, if indicated, outside experts, will examine the project, using the following criteria for evaluation:

- Unequivocal solutions
- Clarity in the statement of the item
- Independence of the questions from each other
- Completeness of the information provided
- Completeness in regard to the steps referenced in the model
- Balance among the item formats
- Life relevance/face validity
- Degree of difficulty appropriate to the target group
- Appropriate expectations in regard to previous knowledge
- Portability to other cultures
- Political correctness
- Lack of group-related bias

Phase 5 ***Informal Pre-Test Using a Sufficiently Large and Balanced Sample Population***

In this phase, information can be obtained (both overall information and for various groups, by gender, educational level, socio-economic status, motivation, previous experience, and language/culture) about the following:

- Consistency
- Reliability
- Difficulty
- Discriminating power of the individual items
- Internal structure of the test

A profound revision is possible on the basis of the pre-test results.

Identifying Critical Item Variables

As described in Part 3, test authors are instructed to take the following item characteristics, among others, into consideration when constructing project tasks:

- Balance among item formats
- Degree of difficulty appropriate to the target group
- Appropriate expectations in regard to previous knowledge

The assessment of these item aspects remained, at least until now, the responsibility of the test authors. Empirically-based variables that determine the difficulty of an item were not identified.

However, first suggestions, including empirical-based ones, for the identification of difficulty-determining variables have now been made. For that purpose, we looked for some typical demands and elements of the project-tasks.

We assume that working on the project-tasks can be described as a cognitive process that includes several solution steps. In the first place, the problem solver has to identify the

demand of the task on the basis of the given information. He or she should extract the conditions which have to be taken into account to find the correct solution. In many cases, working on the task means evaluating given information according to a set of conditions. These conditions, as criteria for evaluation, may or may not be well defined. If they aren't well defined, the problem solver has to interpret the given information and to draw conclusions about the conditions. Depending on the complexity of the task, the evaluation of information requires more or less evaluation steps and different kinds of cognitive processes. In some cases, the problem solver has to take into consideration several connected conditions, and he or she has to distinguish between relevant and superfluous information.

We expect that, in working on the task, the testee will perform some of the following tasks:

- identify some needed information;
- evaluate information according to the given conditions;
- order information according to a given goal;
- integrate information; or
- combine different information and conditions.

It goes without saying that several of these cognitive processes are often required by a single task.

Item variables, which may have an influence on the difficulty and the discriminative power of the items, can be divided into three broad categories:

a) Type of Information
b) Plausibility of Distractors/Complexity of Information
c) Type of Match/Solution Process

The first category, Type of Information, is intended to seek

- The type of information and the number of different types of information used in an item (textual information, quantities, tables, maps, diagrams, architects plans, time tables, schedules)
- The quantity of information, which must be processed
- Quantity of sources of information, as necessary information for the project tasks can potentially be found in different sources (introduction, information folder, item stem)

The second category, Plausibility of Distractors/Complexity of Information, is intended to seek

- The existence of superfluous information
- The criteria for evaluation; are they well defined, ill-defined, or must they be determined by the testee

The third category, Type of Match/Solution Process, takes the following into consideration:

- The main requirements on which the solution of the item is based; these may be analytical reasoning only, a combination of analytical reasoning and practical knowledge, or practical knowledge only.
- The type of processing; a distinction will be made between "identifying," "evaluating," "ordering," "integrating," and "combining."
- The connectivity of cognitive processes: independent, linearly ordered or mutually dependent. Cognitive processes are connected, for example, if the decision on one response alternative depends on the decision on another.
- Meta-level reasoning: whether or not some reasoning about consistency and necessity of information is necessary to solve the item.

- Generation of ideas: if it is required or not.
- The number of criteria which has to be taken into account to decide upon an item.
- The number of evaluation steps needed to solve the item. To find out this number, an idealized solution process that assumes that the testee goes through the item step-by-step in a well-defined manner is described. Each single decision on an element of the item, such as a response alternative, is then counted as one evaluation step.

The importance of the item format as well as the type of question presentation (meaning whether the question is positively or negatively formulated) is also to be analyzed.

With further theoretical and empirical work on critical item variables, the number of potentially relevant variables should be reduced to several central, empirically-confirmed variables.

The first analysis of 64 items, based on some selected item variables, showed that the difficulty of an item is dependent on the question presentation, whether the evaluation criteria are well-defined, and on the number of evaluation steps necessary for solving an item. Items with well-defined evaluation criteria and a low number of necessary evaluation steps are easier to solve.

If one analyzes the discrimination power of the items as dependent on the item variable, one sees that items with many evaluation steps have a lower discrimination power, as do items that require practical knowledge for the solution process.

The fact that purely analytical tasks have a higher discrimination power supports the interpretation of the project approach: an analytical reasoning score as a measure for an essential part of problem solving.

Building an Interpretative Scheme

It is intended to build an interpretative scheme for analytical reasoning in real world contexts scores based on the identification of critical item variables and proficiency levels for score interpretation.

The procedure for identifying proficiency levels will follow Beaton's anchored proficiency procedure (Beaton & Allen, 1992). If a one-dimensional IRT scale has been established, proficiency scaling may be used to define discrete levels of competency. For the description of these levels of competency, results of the critical item analysis will be used as they provide information on the item features which may account for a large percentage of the variance in task difficulty.

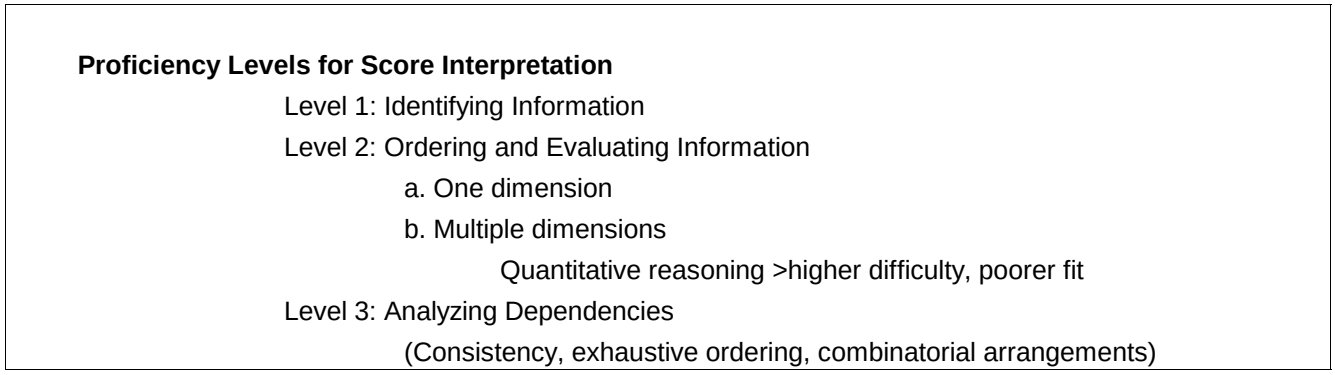
We expect that reporting will be based on a single latent trait, estimated from data for two or more projects. Data from a field test in one of the German States, which employed a problem-solving test for pupils developed according to the project-approach, showed that we might be able to distinguish between three proficiency levels, namely 1) identifying information, 2) ordering and evaluating information, and 3) analyzing dependencies (see Figure 2).

Beyond this, we intend to gather more information for validating the problem-solving score.

The correlations of analytical reasoning in real world contexts with other life skills will be available from the main study. They can contribute to specifying the scope of this score as an essential part of problem solving.

An interview study using the IBF-projects will be conducted by the SRED, a Swiss Research Institute. It will provide information about the solution steps used by the testees and about different solution strategies as they are described by the testees themselves.

Figure 2: Proficiency Levels for Score Interpretation



Authors

Judith Ebach	*Institute of Educational Research, IBF, Bonn, Germany*	Eckhard Klieme	*Max-Planck Institute of Educational Research, Berlin, Germany*
Anne Hensgen	*Institute of Educational Research, IBF, Bonn, Germany*	Eftychia Sidiropoulou	*Institute of Educational Research, IBF, Bonn, Germany*

References

Beaton, A.E. & Allen, N.L. (1992) Interpreting scales through scale anchoring. *Journal of Educational Statistics, 17,* 191-204.

Bloom, B. S. & Broder, J. (1950). *Problem-solving processes of college students.* The University of Chicago Press.

Cobb, P. & Bowers, J. (1999). Cognitive and situated learning. Perspectives in theory and practice. *Educational Researcher, 28,* 4-15.

Ennis, R.H. (1996). *Critical thinking.* Upper Saddle River, N.J. Prentice Hall.

Frensch, P., & Funke, J. (Eds.) (1995). *Complex problem solving*. Mahwah, NJ: Lawrence Erlbaum.

Hensgen, A. & Blum, F. (1998). Prüfen beruflicher Handlungskompetenz: Das KoPrA-Konzept. In J. U. Schmidt (Hrsg.), *Zeitgemäß ausbilden – zeitgemäß prüfen. Berichte zur beruflichen Bildung, Heft 212,* 135-164. Bielefeld: Bertelsmann.

Hensgen, A. & Klieme, E. (1998). Wie bewähren sich die neuen Aufgaben zur Erfassung von Handlungskompetenz in den schriftlichen Zwischen- und Abschlußprüfungen für Bürokaufleute? In J. U. Schmidt (Hrsg.), *Zeitgemäß ausbilden – zeitgemäß prüfen. Berichte zur beruflichen Bildung, Heft 212* , 89-103. Bielefeld: Bertelsmann.

Klieme, E., Funke, J., Leutner, D. & Reimann, P. (1998). Erfassung von Problemlösefähigkeit als fächer-übergreifender Kompetenz (Assessment of problem-solving competence as a cross curricular competence). Unpublished paper, Max Planck Institute for Human Development, Berlin.

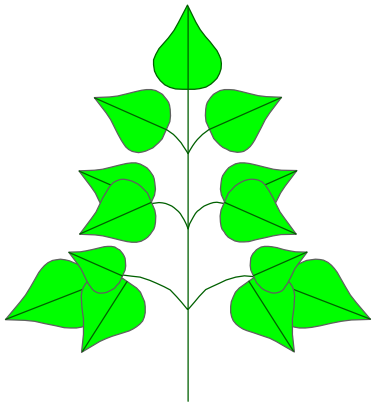
Mayer, R.E., & Wittrock, M.C. (1996). Transfer of problem solving. In D.E. Berlyne (Ed), *Handbook of research on teaching and learning*. New York: Macmillan.

O'Neil, H. (1997). Literature review on problem solving. Paper presented for the International Life Skills Survey (ILSS). Paris.

Sternberg, R.J. & Kaufman J.C. (1998). Human abilities. *Annual Review of Psychology, 49,* 479-502.

Süss, H.M. (1996). *Problemlösen und intelligenz*. Opladen: Westdeutscher Verlag

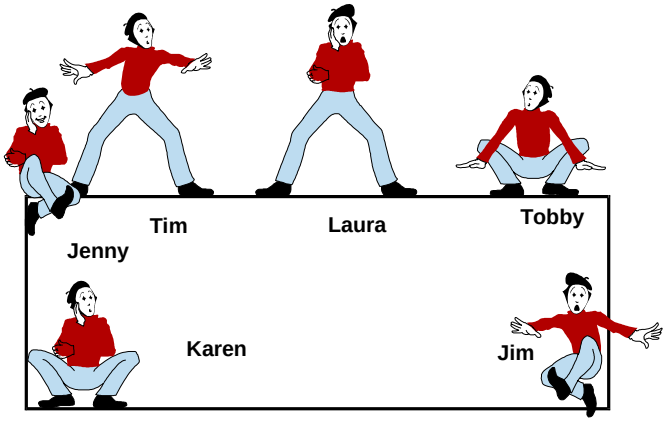
Annex A: Example of a "Project" Tested with German Students



Project: Laying Out a School Garden

Imagine:

Your school can use a neighboring property for free. A soccer field and a school garden will be created. You have established a school garden team with some classmates and now you want to lay out the school garden.



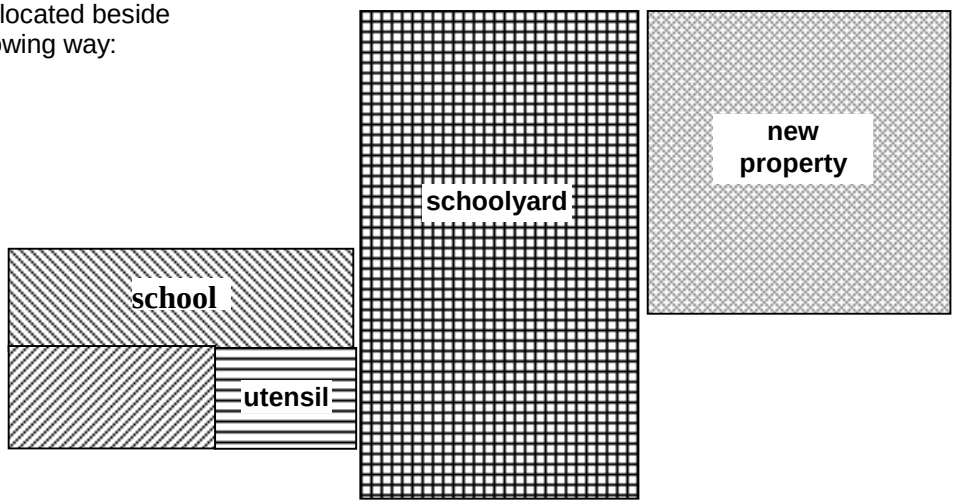
What do you have to do?

You have to

- Take wishes into account
- Collect information
- Organize the work
- Plant the flowers in the flower bed
- Look after the plants

Taking wishes into account:

The new property is located beside the schoolyard in the following way:



Your school-garden team has come to an agreement with the soccer team and you have written down the following requests and guidelines:

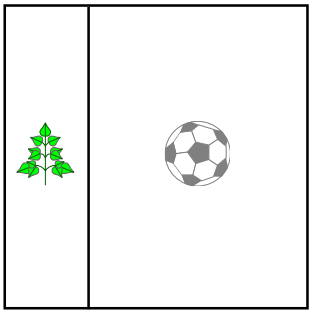
- Only a quarter of the new property should be used for the school garden.
- The school-garden team would like to have the garden entrance as close as possible to the utensil-room.
- The children should not have to pass the garden if they want to go to the soccer field.
- The soccer team would like to have a rectangular space.

Question:

Which of the following partitions can fulfill all requests and guidelines? Circle <u>one</u> answer!

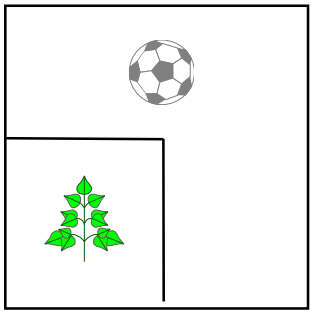
a

schoolyard



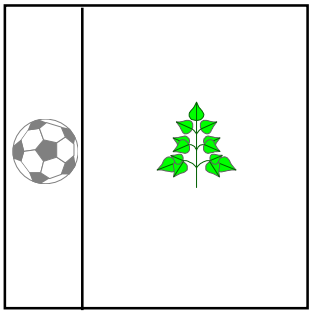
b

schoolyard



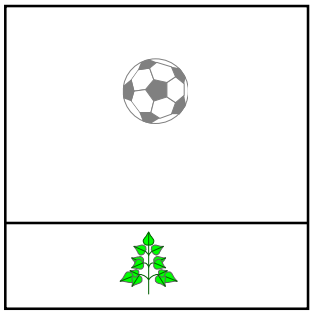
c

schoolyard



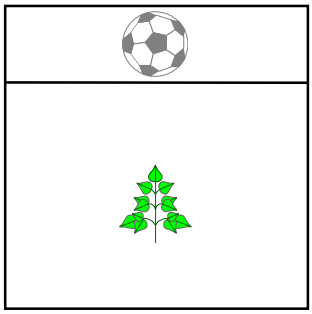
d

schoolyard



e

schoolyard



Collecting Information:

You want to learn about different possibilities for laying out the garden. In the library you are looking for information on the following six topics:

Fence construction	Garden pond
Vegetable bed	Fruit shrubs
Flower garden	Setting of pathways/pavings

You find the following books, but you can only check out three books at one time.

a

The Garden Pond

Tips for planning and cultivating your garden pond

b

Healthy and Delicious From Your Own Garden

All about cultivation of vegetables and fruits

c

The Most Beautiful Flowers From January To October

Information and tips for the arrangement of your flower-garden

d

Self-Made

Tips for woodwork and pavings in the garden: summer houses, fences, pathways, walls, stairways

e

Flowers, Shrubs and Vegetables

How to lay out a useful and beautiful garden

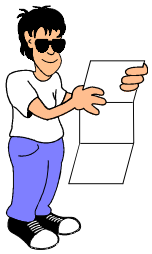
Question: Which books should you take if you want to get information about all six topics? Circle the letters of the <u>three</u> right books!

a b c d e

Organizing the work:

Your team wants to have completed the first steps of gardening by the school party in two weeks. Therefore you need to complete the following tasks:

a. Prepare a bed for the flowers
b. Plant the flowers
c. Sow the vegetables
d. Prepare the new bed with garden mould
e. Buy garden mould and bring it to the school
f. Buy vegetable-seed
g. Plant shrubs



Question: Which sequence of the work-tasks is best? Circle <u>one</u> answer!

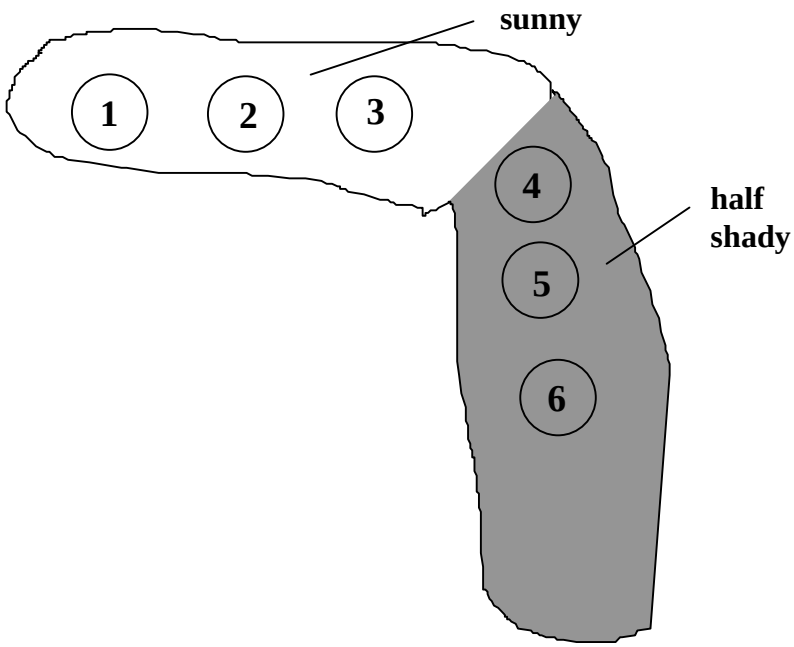
1	a - d - f - g - c - e - b
2	a - f - e - d - c - g - b
3	f - c - e - g - d - a - b
4	a - d - f - c - b - e - g
5	a - e - d - c - f - b - g

Planting the flowers:

For planting the flowers you have to consider:

- Early blooming (March–July) and late blooming flowers (July–October) should alternate.
- In the corner (places 3 and 4) there should be two high-growing flowers.
- Part of the garden is in the shade for several hours of the day.

You have six places where you can plant:



You can use the following flowers that you have picked out in a catalog:

a

Bell-flower

Location: half-shady -shady
Blossom: May–June
Height: 15 cm

b

Fluffy-flower

Location: sunny - half-shady
Blossom: July–August
Height: 120 cm

c

Pink

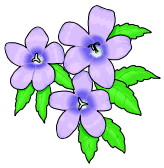
Location: sunny
Blossom: May–June
Height: 25–30 cm

d

Red foxglove

Location: sunny
Blossom: June–July
Height: 80–120 cm

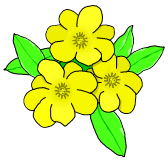
e



Violet

Location: half-shady
Blossom: July–September
Height: 10–15 cm

f



Primrose

Location: sunny
Blossom: July–September
Height: 10–20 cm

Question: Which flower has to be planted in which spot? Circle <u>one</u> answer in each case!

1. On place 1 has to be planted flower	a	b	c	d	e	f
2. On place 2 has to be planted flower	a	b	c	d	e	f
3. On place 3 has to be planted flower	a	b	c	d	e	f
4. On place 4 has to be planted flower	a	b	c	d	e	f
5. On place 5 has to be planted flower	a	b	c	d	e	f
6. On place 6 has to be planted flower	a	b	c	d	e	f

Looking after the plants:

For the summer vacation you made a plan of who would water the garden and when. Unfortunately, some plants still dried up during the summer. Now you want to know how this happened.
Your plan:

1st Vacation-week	
5 July:	Tim
6 July:	
7 July:	Jim
8 July:	
9 July:	Jim
10 July:	
11 July:	Laura

2nd Vacation-week	
12 July:	Karen
13 July:	
14 July:	Karen
15 July:	
16 July:	Karen
17 July:	
18 July:	Laura

3rd Vacation-week	
19 July:	Jenny
20 July:	
21 July:	Jenny
22 July:	
23 July:	Karen
24 July:	
25 July:	Karen

4th Vacation-week	
26 July:	Jim
27 July:	
28 July:	Jim
29 July:	
30 July:	Jim
31 July:	
1 August:	Laura

5th Vacation-week	
2 August:	Laura
3 August:	
4 August:	Tobby
5 August:	
6 August:	Tobby
7 August:	
8 August:	Tobby

6th Vacation-week	
9 August:	Jenny
10 August:	
11 August:	Tim
12 August:	
13 August:	Tim
14 August:	
15 August:	Tim

- Karen was ill from 24 July until 27 July.
- Tobby forgot to water the plants.
- Jim went on vacation unexpectedly on 22 July for a week.
- It rarely rained during the summer, but in the fifth week, there were several violent thunderstorms.

Question: **When and why did your flowers dry up? Circle <u>one</u> statement!**

a	In the fourth week, because Jim was on vacation and Karen could not water the plants more than once.
b	In the second week, because Karen was ill.
c	In the fifth week, because Tobby forgot to water the plants.
d	You are not sure exactly when they dried up, but they dried up because you don't have much experience yet with plants and everyone watered them too little.

Annex B: Realized Projects for the Feasibility Study

- **Renovating a clubhouse** (separate information folder)
  Questions, for example, on what should be renovated, the coordination of work, comparison of costs of materials, and possible reasons for bad work.

- **Planning a trip and a family reunion** (separate information folder)
  Questions, for example, on a possible date for the reunion, the advantages of different modes of transportation, the flight number to choose, the steps to be taken before booking the flight, and the control of the flight ticket.

- **Setting up a space station**
  Questions, for example, on the goals that can be better reached in space than on earth, the requirements for the station, the experts needed for different work-steps, the sequence of the work, the selection of astronauts, and the correctness of the transportation plan.

- **Buying a bicycle**
  Questions, for example, on where to get information about kinds of bicycles, comparison of different bicycles in regard to the requirements, and the decision on where to buy the bicycle.

- **Organizing the visit of a foreign chorus** (separate information folder)
  Questions, for example, on the wishes of the visitors, the proper sequence of work-steps to be taken, the composition of a program for the visit, and the budget for the visit.

- **Organizing a sports meeting**
  Questions, for example, on possible sequence of events, location of facilities, necessary entrance fees, possible reactions to changes, and the causes for problems.

- **Looking for a new job**
  Questions, for example, on the disadvantages of the current job, analysis of positions offered, and plans for what has to be done before starting the new job.

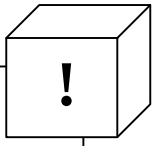
- **Looking for a new apartment and moving**
  Questions, for example, on the advantages of different lodging types, the comparison of different lodgings in regard to the requirements and the price, the organization of transport, the shortest route to the new house, and the mistakes in the contract.

Annex C: Examples of Items Referring to Different Steps with Different Item Formats

Step: Define the Goals

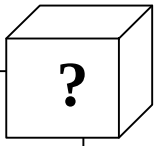
General question: Which goals or requests come into conflict with each other or cannot be reached simultaneously?

Item format: Multiple-choice Item



First you have gathered the club members' suggestions:

a) refurbish the wooden floor,
b) paint the exterior of the building,
c) add a restroom with a sink,
d) lay new ceramic tile everywhere,
e) build a bar,
f) beautify the grounds,
g) leave the floor plan of the house as it is,
h) only renovate the interior of the clubhouse.



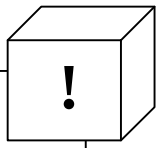
Some of the suggestions are contradictory. There is, however, one suggestion that is compatible with all the others. Which one is it?
Please select one suggestion!

a **b** **c** **d** **e** **f** **g** **h**

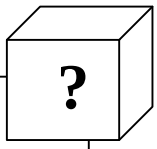
Step: Analyze the Situation

General question: Which information do you need?

Item format: Free-Response Item



The family reunion is scheduled to begin at 10 a.m. at a hotel in the hiking area. Because you want to bring gifts for your relatives and you do not want to wear your hiking clothes while traveling, you must check in a suitcase. Now you want to determine if you can arrive on the day of the reunion, July 12, although you do not want to leave your house before 6 a.m. that morning.

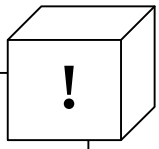


In order to make this decision, what time factors (such as duration etc.) must you take into consideration? ***This information can be found in the information booklet.***

Step: Plan the Solution

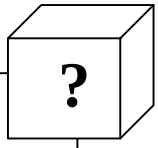
General question: In which sequence must the work-steps be executed?

Item format: Ordering Item



The initial plans are complete. The separate rooms (modules) of the station shall be constructed and completely furnished on earth. Subsequently, they should be brought into orbit and assembled together there. You have made a note of some of the most important work stages.

A) Construction of the separate modules on earth,
B) Setting up the laboratory and the other rooms,
C) Obtain (or commission) the production of the material for the space station construction
D) Develop computer programs for navigating the station,
E) Construction or provision of vehicles for transportation in space,
F) Assembling the modules in space,
G) Training the astronauts and space station crew,
H) Transporting the modules into orbit.

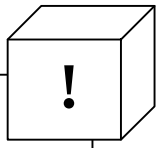


There are different possibilities of the sequence in which these steps can be carried out. Which of the steps listed below is not or are <u>not</u> feasible?
Please select <u>all</u> sequences that are <u>not</u> feasible!

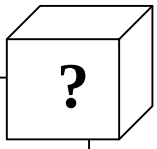
a	C - E - A - D - G - B - H - F
b	G - D - C - A - B - E - H - F
c	E - A - C - F - B - G - D - H
d	D - E - C - A - G - F - H - B

Step: Execute the Plan

Item format: Short response item



One your relatives is unable to meet the rest of you at 10 a.m. as planned. Because of this the meeting time for the reunion has been change to 11 a.m. Consequently, you have decided to arrive on the same day (July 12) as the reunion.



Which flight must you book in order to be punctual, but not too early at the designated meeting place (hotel)?
Please note the flight number!

flight number: ☐

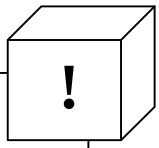
What time do you have to leave your house in order to catch the shuttle bus to the airport?
Please note the correct time!

☐ **o'clock**

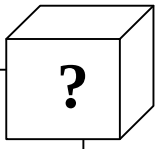
Step: Evaluate the Results

General question: Which causes are responsible for the mistake?

Item format: Correspondence Item



You have completed the first part of the renovations (windows, walls, ceiling, and floors). Now you are all sitting together, relaxing, and contemplating the work that you have done. Some of the things did not work out as you had imagined they would. You think about what could have been possible reasons for this in order to avoid them in the future. The cause of the mistakes could possibly be attributed to the organization of the work, the choice or quality of the materials, and/or the technical skills of the club members.



In which areas could possible reasons for the following mistakes be found?
Please select <u>all</u> possible sources of mistakes:

	Organization	**Materials / tools (equipment)**	**technical skills**
1. The dripless color dripped.	**a**	**b**	**c**
2. Mrs. X. was on vacation at the time when she should have picked the sanding machine.	**a**	**b**	**c**
3. The wooden floor varnish was not dry when Mr. D. wanted to hang the curtains back up.	**a**	**b**	**c**
4. There are visible patches of unevenness on some parts of the floor.	**a**	**b**	**c**
5. The ceiling is lightly smudged.	**a**	**b**	**c**